(2013), **xx**, x, *pp.* 1–14

A penalized algorithm for event-specific rate models for recurrent events

By O. Bouaziz
MAP5, UMR CNRS 8145 and University Paris Descartes, Paris, France,
olivier.bouaziz@parisdescartes.fr

and A. Guilloux
LSTA, University Pierre et Marie Curie, 4 place Jussieu, Paris, France,
agathe.guilloux@upmc.fr

Summary

We introduce a covariate-specific total variation penalty in two semiparametric models for the rate function of recurrent event process. The two models are a stratified Cox model, introduced in Prentice et al. (1981), and a stratified Aalen's additive model. We show the consistency and asymptotic normality of our penalized estimators. We demonstrate, through a simulation study, that our estimators outperform classical estimators for small to moderate sample sizes. Finally an application to the bladder tumour data of Byar (1980) is presented.

Some key words: Recurrent events process; total variation penalization; Aalen model; Cox model.

1. Introduction

Recurrent events are frequent in clinical or epidemiological studies when each subject experiences repeated events over the time. Standard medical examples include the repetition of asthma attacks, epileptic seizures or tumour recurrences for individual patients. In this context, proportional hazards models have been largely studied in the literature to model the rate or mean functions of recurrent event data. For instance, Andersen & Gill (1982) introduce a conditional Cox model where the recurrent events process is assumed to be a Poisson process. Without this assumption, similar proportional hazards models and extensions are considered in Lawless & Nadeau (1995), Lin et al. (1998), Lin et al. (2000) and Cai & Schaubel (2004).

To model rate functions in a recurrent events context, a different approach consists in fitting a Cox model for any different recurrence. Along these lines, Prentice et al. (1981) introduce two stratified proportional hazards models with event-specifics baseline hazards and regression coefficients. Gap times and conditional models are presented in their paper and a marginal event-specific model is studied in Wei et al. (1989). We refer to Kelly & Lim (2000) for a complete review of existing Cox-based recurrent event models.

Additive models provide an useful alternative to proportional hazards models. For classical counting processes, the Aalen model was first introduced in Aalen (1980) and is extensively studied in McKeague (1988), Huffer & McKeague (1991), Lin & Ying (1994). It is considered in the context of recurrent events in Scheike (2002). We propose in this

paper to consider an event-stratified version of the Aalen model, in the manner of Prentice et al. (1981).

As demonstrated in the following, event-stratified models allow more flexibility but suffer from over-parametrization as soon as the sample size is not large enough with respect to the number of covariates and the number of recurrent events. We address this drawback by introducing new estimators defined as minimizers of penalized empirical risks. More specifically, we consider a covariate-specific total variation penalty.

The remainder of this article is organized as follows. The multiplicative and additive models studied in this paper are presented in Section 1. In Paragraph 2·4, we describe our novel algorithms. It requires preliminary details on inference in these two models, which are given in Paragraphs 2·2 and 2·3. Consistency and asymptotics normality of the estimators are derived in Section 3. Simulation studies and a real data analysis are provided in Sections 4 and 5. A discussion and some concluding remarks are contained in Section 6.

2. Models and algorithm

2·1. *Models*

Let D denote the time of the terminal event and $N^*(t)$ the number of recurrent events before time t. The end-point of the observation is $\tau > 0$. The p-dimensional process of covariates is denoted by X and ρ_0 represent the rate function. The event-specific rate function of the process N^* is then defined as

$$\mathbb{E}\big(dN^*(t) \mid X(t), D \geq t, N^*(t) = s - 1\big) = \mathbf{1}(D \geq t)\rho_0(t, s, X(t))dt,$$

for t in $[0, \tau]$ and $s = 1, \ldots, B$. Apart from the stratification, this definition of the rate function can be found in Scheike (2002).

We consider two semiparametric models for the function ρ_0. The first one is an event-specific multiplicative rate model introduced in Prentice et al. (1981). In this model, the rate function is specified, for t in $[0, \tau]$, by

$$\rho_0(t, s, X(t)) = \alpha_0(t, s) \exp\left(X(t)\beta_0(s)\right) \tag{1}$$

where for each event number s, $\beta_0(s)$ is an unknown p-dimensional vector of parameters and α_0 is an unknown baseline function.

Following Scheike (2002), and Zeng & Cai (2010), we also propose to consider its additive counterpart. The rate function in our event-specific additive model is then for t in $[0, \tau]$:

$$\rho_0(t, s, X(t)) = \left(\alpha_0(t, s) + X(t)\beta_0(s)\right). \tag{2}$$

The models, where β_0 is constant over the events are refereed to as constant models in what follows.

We consider the problem of estimating the unknown parameter β_0, in stratified models (1) and (2) on the basis of data from n independent and identically distributed random variables. Introduce the censoring time C. In a random sample of n subjects, the data consist of $\{N_i(t), T_i, \delta_i, X_i(t), t \leq \tau\}$, $i = 1, \ldots, n$ where $N_i(t) = N_i^*(t \wedge C_i)$, $T_i = D_i \wedge C_i$ is the minimum between D_i and C_i, $\delta_i = \mathbf{1}(D_i \leq C_i)$ and $(X_i(t), 0 \leq t \leq T_i)$ is the covariates process. The next assumption characterizes the dependence mechanism between the censoring time and the other variables.

Assumption 1. For all $s = 1, \dots, B$ and t in $[0, \tau]$,

$$\mathbb{E}\big(dN^*(t) \mid X(t), D \wedge C \geq t, N^*(t) = s-1\big) = \mathbb{E}\big(dN^*(t) \mid X(t), D \geq t, N^*(t) = s-1\big).$$

Note that this assumption is slightly weaker than assuming the independence between C and (N^*, D, X). A similar assumption can be found for instance in Lin et al. (2000). We also impose the following conditions on the tails of the distribution of T and N.

Assumption 2. There exists a nonnegative integer B such that

(i) $\forall t \in [0, \tau]$, $\mathbb{P}\big(N(t) \leq B\big) = 1$,
(ii) $\forall t \in [0, \tau]$, $\forall s = 1, \dots, B$, $\mathbb{P}\big(T \geq t, N(t) = s-1 \mid X(t)\big) > 0$.

Assumption 2 (i) ensures that in models (1) and (2), the total number of observed events is almost surely bounded. It is standard for inference for recurrent events process, see e.g. Dauxois & Sencey (2009), Scheike (2002) or Bouaziz et al. (2013).

Under Assumption 2, the unknown vector of parameters β_0 has $p \times B$ unknown coefficients to be estimated. For reasonable sizes of sample n, these models are over-parametrized in the sense that, when $\sqrt{n} \leq p \times B$, the estimators show very poor behaviours (see Section 4 for an illustration). On the other hand, simpler forms of models (1) and (2), in which the unknown parameter does not change with the event, $\beta_0(s) = \beta_0$, might be too poor to accurately fit the data (see also Section 4 and the discussion in Kelly & Lim (2000)). In this paper, we aim at providing estimators realizing a compromise between these two situations.

In the following, we define, for each individual i, the event-specific at-risk function Y_i^s and the overall at-risk function Y_i for all t in $[0, \tau]$:

$$Y_i^s(t) = \mathbf{1}(T_i \geq t, N_i(t) = s), \quad Y_i(t) = \sum_{s=1}^{B} Y_i^s(t) = \mathbf{1}(T_i \geq t).$$

2·2. *Inference in the multiplicative model*

As in Prentice et al. (1981), in the multiplicative event-specific model (1), an estimator $\hat{\beta}_{ES/mult}$ of the unknown parameter $\beta_0 \in \mathbb{R}^{p \times B}$ is defined as the maximizer of the partial log-likelihood, or equivalently as

$$\begin{aligned}\hat{\beta}_{ES/mult} &\in \underset{\beta \in \mathbb{R}^{p \times B}}{\operatorname{argmin}}\, L_n^{PL}(\beta) \qquad (3)\\ &= \underset{\beta \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} \left[-\frac{1}{n} \sum_{s=1}^{B} \sum_{i=1}^{n} \int \left\{ X_i(t)\beta(s) - \log\left(\sum_{j=1}^{n} Y_j^s(t) \exp\left(X_j(t)\beta(s)\right) \right) \right\} Y_i^s(t) dN_i(t) \right].\end{aligned}$$

An estimator $\hat{\beta}_{C/mult}$ in the constant model is defined as

$$\hat{\beta}_{C/mult} \in \underset{\beta \in \mathbb{R}^{p}}{\operatorname{argmin}} \left[-\frac{1}{n} \sum_{i=1}^{n} \int \left\{ X_i(t)\beta - \log\left(\sum_{j=1}^{n} Y_j(t) \exp\left(X_j(t)\beta\right) \right) \right\} Y_i(t) dN_i(t) \right]. \qquad (4)$$

2·3. *Inference in the additive model*

As noticed in Martinussen & Scheike (2009a,b) or Gaiffas & Guilloux (2012), in the usual additive hazards model, the estimator $\hat{\beta}_{ES/add}$ of the unknown parameter $\beta_0 \in$

$\mathbb{R}^{p\times B}$ can be written as the minimizer of a (partial) least-squares criterion:

$$\hat{\beta}_{ES/add} \in \underset{\beta\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\, L_n^{PLS}(\beta) = \underset{\beta\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \sum_{s=1}^{B} \left\{\beta(s)^\top \mathbf{H}_n(s)\beta(s) - 2\boldsymbol{h}_n(s)\beta(s)\right\}, \quad (5)$$

where for all $s \in \{1,\ldots,B\}$, $\mathbf{H}_n(s)$ are $p\times p$ symetrical positive semidefinite matrices equal to

$$\frac{1}{n}\sum_{i=1}^{n}\int Y_i^s(t)\Big(X_i(t) - \bar{X}^s(t)\Big)^{\otimes 2} dt,$$

and where $\boldsymbol{h}_n(s)$ are p-dimensional vectors equal to

$$\frac{1}{n}\sum_{i=1}^{n}\int \mathbf{1}(N_i(t)=s)\Big(X_i(t) - \bar{X}^s(t)\Big) dN_i(t),$$

with $\bar{X}^s(t) = \sum_{i=1}^n X_i(t)Y_i^s(t)/\sum_{i=1}^n Y_i^s(t)$. We show in the Appendix why this criterion is a relevant strategy in the additive event-specific model.

On the other hand, an estimator $\hat{\beta}_{C/add}$ in the constant model is defined as

$$\hat{\beta}_{C/add} \in \underset{\beta\in\mathbb{R}^{p}}{\operatorname{argmin}} \Big(\beta^\top \mathbf{H}_n\beta - 2\boldsymbol{h}_n\beta\Big), \text{ with } \mathbf{H}_n = \sum_{s=1}^{B}\mathbf{H}_n(s) \text{ and } \boldsymbol{h}_n = \sum_{s=1}^{B}\boldsymbol{h}_n(s). \quad (6)$$

2·4. *A total-variation penalty*

To overcome the possible over-parametrization of models (1) and (2), we propose to define penalized versions of criteria (3) and (5). For all $\beta = (\beta(s), s=1,\ldots,B)$ with $\beta(s) = (\beta^1(s),\ldots,\beta^p(s))$, define for all $j=1,\ldots,p$

$$\beta^j = (\beta^j(1),\ldots,\beta^j(B)) \text{ and } \mathrm{TV}(\beta^j) = \sum_{s=2}^{B}|\beta^j(s) - \beta^j(s-1)| = \sum_{s=2}^{B}|\Delta\beta^j(s)|. \quad (7)$$

We now consider the minimizers of the partial log-likelihood (respectively the partial least-squares) penalized with a covariate specific total variation. Define the penalized estimators in models (1) and (2) as:

$$\hat{\beta}_{\mathrm{TV}/mult} \in \underset{\beta\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \left\{L_n^{PL}(\beta) + \frac{\lambda_n}{n}\sum_{j=1}^{p}\mathrm{TV}(\beta^j)\right\} \text{ and} \quad (8)$$

$$\hat{\beta}_{\mathrm{TV}/add} \in \underset{\beta\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \left\{L_n^{PLS}(\beta) + \frac{\lambda_n}{n}\sum_{j=1}^{p}\mathrm{TV}(\beta^j)\right\}. \quad (9)$$

These penalized algorithms can be rewritten as lasso algorithms (the details are given in Supplementary Material).

3. Asymptotic results

We successively provide the asymptotic results for the estimators $\hat{\beta}_{\mathrm{TV}/add}$ in the additive model and $\hat{\beta}_{\mathrm{TV}/mult}$ in the multiplicative model. In both models, the following condition is mandatory.

Assumption 3. The covariates process $X(\cdot)$ is of bounded variation on $[0,\tau]$.

Define for all $s=1,\dots,B$ the centered process $M^s(t)=N(t)-\mathbb{E}\big(N(t)\mid X(t), D\wedge C\geq t, N(t)=s-1\big)$ and the $p\times p$ matrix

$$\mathbf{H}(s) := \int \mathbb{E}[Y^s(t)X(t)^\top X(t)]dt - \int \frac{(\mathbb{E}[Y^s(t)X(t)])^{\otimes 2}}{\mathbb{E}[Y^s(t)]}dt,$$

which from Assumption 2 (ii) is well defined.

Theorem 1. *Assume that, for each $s=1,\dots,B$, $\mathbf{H}(s)$ is non-singular and that Assumptions 1, 2 and 3 are fulfilled.*

1. If $\lambda_n/n\to 0$ as $n\to\infty$ then $\hat{\beta}_{\mathrm{TV}/add}$ converges to β_0 in probability.

2. If $\lambda_n/\sqrt{n}\to\lambda_0\geq 0$ as $n\to\infty$ then $\sqrt{n}(\hat{\beta}_{\mathrm{TV}/add}-\beta_0)$ converges in distribution to

$$\begin{aligned}\underset{u\in\mathbb{R}^p}{argmin}\,\Lambda_{add}(u) = \underset{u\in\mathbb{R}^p}{argmin}\Big[&\sum_{s=1}^{B}\Big\{u(s)^\top\mathbf{H}(s)u(s)-2u(s)^\top\xi_{add}(s)\Big\}\\ &+\lambda_0\sum_{j=1}^{p}\sum_{s=2}^{B}\big\{|\Delta u^j(s)|\mathbf{1}(\Delta\beta^j(s)=0)+\,sgn(\Delta\beta^j(s))(\Delta u^j(s))\mathbf{1}(\Delta\beta^j(s)\neq 0)\big\}\Big],\end{aligned}$$

and for each s, $\xi_{add}(s)$ is a centered p-dimensional gaussian vector with covariance matrix equal to

$$\mathbb{E}\left[\Big(\int_0^\tau (X(t)-\mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])\mathbf{1}(N(t)=s)dM^s(t)\Big)^{\otimes 2}\right].$$

Define for all $s=1,\dots,B$ and for all $t\in[0,\tau]$,

$$s^{(l)}(s,t,\beta)=\mathbb{E}[Y^s(t)X(t)^{\otimes l}\exp(X(t)\beta(s))], l=0,1,2.$$

Introduce $\mathbf{e}(s,t,\beta)=s^{(1)}(s,t,\beta)/s^{(0)}(s,t,\beta)$, $\mathbf{v}(s,t,\beta)=s^{(2)}(s,t,\beta)/s^{(0)}(s,t,\beta)-\mathbf{e}(s,t,\beta)^{\otimes 2}$ and $\mathbf{\Sigma}(s,\beta)=\int\mathbf{v}(s,t,\beta)\mathbb{E}[Y^s(t)dN(t)]$. For any $s=1,\dots,B$ and for any $t\in[0,\tau]$, the three functions $s^{(l)}(s,t,\beta_0)$ are bounded from Assumption 3 and $\mathbf{e}(s,t,\beta),\mathbf{v}(s,t,\beta)$ and $\mathbf{\Sigma}(s,\beta)$ are finite from Assumptions 2 and 3.

Theorem 2. *Assume that for each $s=1,\dots,B$, $\mathbf{\Sigma}(s,\beta_0)$ is non-singular and that Assumptions 1, 2 and 3 are fulfilled.*

1. If $\lambda_n/n\to 0$ as $n\to\infty$ then $\hat{\beta}_{\mathrm{TV}/mult}$ converges to β_0 in probability.

2. If $\lambda_n/\sqrt{n}\to\lambda_0\geq 0$ as $n\to\infty$ then $\sqrt{n}(\hat{\beta}_{\mathrm{TV}/mult}-\beta_0)$ converges in distribution to

$$\begin{aligned}\underset{u\in\mathbb{R}^p}{argmin}\,\Lambda_{mult}(u) = \underset{u\in\mathbb{R}^p}{argmin}\Big[&\sum_{s=1}^{B}\Big\{\frac{1}{2}u(s)^\top\mathbf{\Sigma}(s,t,\beta_0)u(s)+u(s)^\top\xi_{mult}(s)\Big\}\\ &+\lambda_0\sum_{j=1}^{p}\sum_{s=2}^{B}\Big\{|\Delta u^j(s)|\mathbf{1}(\Delta\beta_0^j(s)=0)+sgn(\Delta\beta_0^j(s))(\Delta u^j(s))\mathbf{1}(\Delta\beta_0^j(s)\neq 0)\Big\}\Big],\end{aligned}$$

and for each s, $\xi_{mult}(s)$ is a centered p-dimensional gaussian vector with covariance matrix equal to

$$\mathbb{E}\left[\left(\int_0^\tau (X(t) - \mathbf{e}(s,t,\beta_0))\, Y^s(t) dM^s(t)\right)^{\otimes 2}\right].$$

Theorems 1 and 2 prove the consistency and asymptotic normality of our estimators (8) and (9). This assures that they behave better than the constant estimators when β_0 is non constant. In addition, the considered penalty will induce sparsity for each covariate $j = 1, \dots, p$ in the successive differences $\Delta\beta^j(s),\ s = 1, \dots, B$. As a consequence, the effects of a covariate on two consecutive events will often be equal. We show, in the following simulation study, that this induced sparsity ameliorates the behaviour of our estimators compared to the unconstrained ones (defined in Equations (3) and (5)).

4. Simulation studies

We compare the performances of the penalized estimators (8) and (9), the constant ones (4) and (6) , and the unconstrained ones (3) and (5). To mimic the bladder tumour cancer dataset studied in Section 5, we set $p = 4$ and consider $B = 5$ recurrent events for the estimation. In the multiplicative and additive models, the sample size n varies from $n = 50 = 2{\cdot}5\, pB$ to $n = 1000 \simeq (pB)^{2\cdot 3}$.

We draw the $p = 4$ covariates from uniform distributions and set the parameters values at $\beta_0^1 = (0, 0, b_1, b_1, 0, \dots, 0)$, $\beta_0^2 = (b_2, \dots, b_2)$, $\beta_0^3 = b_3(1, 2, 3, \dots)$ and $\beta_0^4 = (0, \dots, 0)$. We generate recurrent event times from the multiplicative (1) and additive (2) models with baseline defined through the Weibull distribution with shape parameter $a_{\mathcal{W}}$ and scale parameter 1. The death and censoring times are generated from exponential distributions with parameters a_D and a_C respectively. We set the value of parameter $a_{\mathcal{W}}$ at $2{\cdot}5$. Finally, the values of a_D and a_C are empirically determined to obtain $\mathrm{P}_{\mathrm{obs}} = 28 - 29\%$ and $14 - 15\%$ of individuals experiencing the fifth event.

To evaluate the performances of the different estimators, we conduct a Monte Carlo study with $M = 200$ experiences. The estimation accuracy is investigated for each method via a mean squared rescaled error defined as

$$\text{MSE} = \frac{1}{M} \sum_{m=1}^{M} \frac{\|\hat{\beta}_m - \beta_0\|^2}{\|\beta_0\|^2}, \tag{10}$$

where $\hat{\beta}_m$ is the estimation in the sample m. We furthermore study the detection power of non-constant (respectively constant) covariate effects by computing mean false positive (FP) rates and mean false negative (FP) rates for each method. They are defined, for an estimation $\hat{\beta}_m$, as

$$\text{FP}(\hat{\beta}_m) = \text{Card}\left(j \in \{1, \dots, p\} \text{ s.t. } \text{TV}(\hat{\beta}^j) \neq 0 \text{ and } \text{TV}(\beta_0^j) = 0\right) \tag{11}$$

and

$$\text{FN}(\hat{\beta}_m) = \text{Card}\left(j \in \{1, \dots, p\} \text{ s.t. } \text{TV}(\hat{\beta}^j) = 0 \text{ and } \text{TV}(\beta_0^j) \neq 0\right), \tag{12}$$

where TV is defined in (7).

As expected, the constant model is biased and behave poorly for our choice of a non-constant β_0. The comparison between the unconstrained and penalized estimators is

Table 1. *Simulation results in the multiplicative model for* $\mathrm{P}_{obs} = 28\%$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	0·100	2	0	0·412	0	2	0·054	1·44	0·03	0·044	0·82	0·02
100	0·030	2	0	0·415	0	2	0·025	1·54	0	0·019	0·76	0
500	0·006	2	0	0·413	0	2	0·008	1·76	0	0·006	0·30	0
1000	0·005	2	0	0·415	0	2	0·006	1·81	0	0·006	0·05	0

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 2. *Simulation results in the multiplicative model for* $\mathrm{P}_{obs} = 14\%$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0·440	0	2	0·161	1·37	0·185	0·137	0·82	0·19
100	0·566	2	0	0·434	0	2	0·053	1·55	0·005	0·042	0·88	0
500	0·014	2	0	0·433	0	2	0·016	1·84	0	0·012	1·06	0
1000	0·009	2	0	0·433	0	2	0·011	1·89	0	0·010	0·68	0

MSE: mean squared error, FP: false positives, FN: false negatives, NA: non applicable .

Table 3. *Simulation results in the additive model for* $\mathrm{P}_{obs} \simeq 28\%$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	4·986	2	0	0·416	0	2	0·467	0·98	0·58	1·142	0·65	0·81
100	0·935	2	0	0·351	0	2	0·254	1·38	0·21	0·353	0·86	0·48
500	0·135	2	0	0·309	0	2	0·079	1·91	0·01	0·094	1·44	0·08
1000	0·071	2	0	0·299	0	2	0·049	1·98	0	0·05	1·64	0

MSE: mean squared error, FP: false positives, FN: false negatives

Table 4. *Simulation results in the additive model for* $\mathrm{P}_{obs} \simeq 14\%$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0·505	0	2	0·781	0·95	0·81	2·368	0·86	0·97
100	4·114	2	0	0·393	0	2	0·707	1·450	0·27	0·84	1·11	0·52
500	0·339	2	0	0·330	0	2	0·154	1·975	0·01	0·19	1·67	0·06
1000	0·171	2	0	0·320	0	2	0·097	1·995	0	0·12	1·80	0·02

MSE: mean squared error, FP: false positives, FN: false negatives, NA: non applicable .

in favour of our estimator in all four cases as long as n is smaller than p^2. When the percentage of individuals experiencing the fifth event drops, non-constant estimators are slightly less accurate. Algorithms are not able to compute all $M = 200$ unconstrained estimators for $n = 50$. For $p = 4$, $B = 5$, $n = 100$ and $\mathrm{P}_{\mathrm{obs}} = 14\%$ (which are values close to those encountered in the bladder tumour cancer dataset studied in the next section) our penalized estimators are respectively 5·8, in the additive model, and 10·6, in the multiplicative model, times better than the unconstrained ones in terms of estimation error.

Surprisingly the number of false positives detected by our penalized estimators increases when the sample size increases. A possible solution to ameliorate the latter is to apply the reweighed lasso, or two-steps lasso, as proposed in Candès et al. (2008) (details are given in Supplementary Material). We compute the mean squared error, false positive

and negative rates of the resulting estimator. It shows better false positive rates than the first step penalized estimator, greater false negative rates and comparable mean squared errors.

We repeat the simulation study for $a_{\mathcal{W}} = 2{\cdot}5$ and then for a Gompertz baseline with shape parameter $a_{\mathcal{G}} = 0{\cdot}5$ (and $a_{\mathcal{G}} = 0{\cdot}5$) and scale parameter 1. The results are reported in Supplementary Material. Conclusions are similar.

5. Bladder tumour data analysis

In this section we illustrate the behaviour of our estimators on the bladder tumour cancer data of Byar (1980). These data were obtained from a clinical trial conducted by the Veterans Administration Co-operative Urological Group. One hundred and sixteen patients were randomised to one of three treatments: placebo, pyridoxine or thiotepa. For each patient, the time of recurrence tumours were recorded until the death or censoring times. The number of recurrences ranges from 0 to 10. On the $n = 116$ patients, since $13{\cdot}79\%$ experienced at least five tumour recurrences and only $6{\cdot}9\%$ patients experienced six tumour recurrences or more, we set the parameter B to 5. In addition to the two treatment variables, pyridoxine and thiotepa, two supplementary covariates were recorded for each patient: the number of initial tumours and the size of the largest initial tumour.

Figure 1 displays the estimations obtained from the constant, unconstrained and total variation estimators in the multiplicative model. In order to enforce the variables selection performance of the total variation estimator, the coefficients were estimated using the reweighed lasso. The unconstrained estimator shows very strong variations and is difficult to interpret as such. On the other hand, the constant estimator gives valuable information on the impact of each covariate, but in turn cannot detect a change in variation. Our total-variation estimator reaches compromise: it is not constant but easily interpretable.

For instance, a remarkable aspect of the pyrodixine treatment can be highlighted from the total variation estimation: this treatment produces a protective effect for the first three tumour recurrences but the odds of further recurrences are increased by this treatment. In the same way, an increase in the effect of the initial number of tumours on recurrences is observed from the third recurrence. On the opposite, the effects of the thiotepa treatment or the size of the largest tumour are shown to be constant in the total variation model, the parameter estimates having values similar to the ones obtained in the constant model.

Our conclusions on the treatments effects are in agreement with previous studies on bladder tumours recurrences. For instance, no difference in the rate or time to tumour recurrence was found from patients using pyrodixine with patients using placebo in Tanaka et al. (2011) and Goossens et al. (2012). Moreover, Huang & Chen (2003) and Sun et al. (2006) have respectively studied gap time recurrences in the multiplicative and additive models. The results obtained from the former showed a small protective effect of this treatment while the latter concluded that gap times did not seem related to pyridoxine. These examples illustrate the nice features of our total-variation estimator: it provides sharper results, giving relevant informations on covariates effect with respect to the number of recurrent events experienced by a subject and it provides the ability to detect a change of variation. Further details are provided in Supplementary Material.

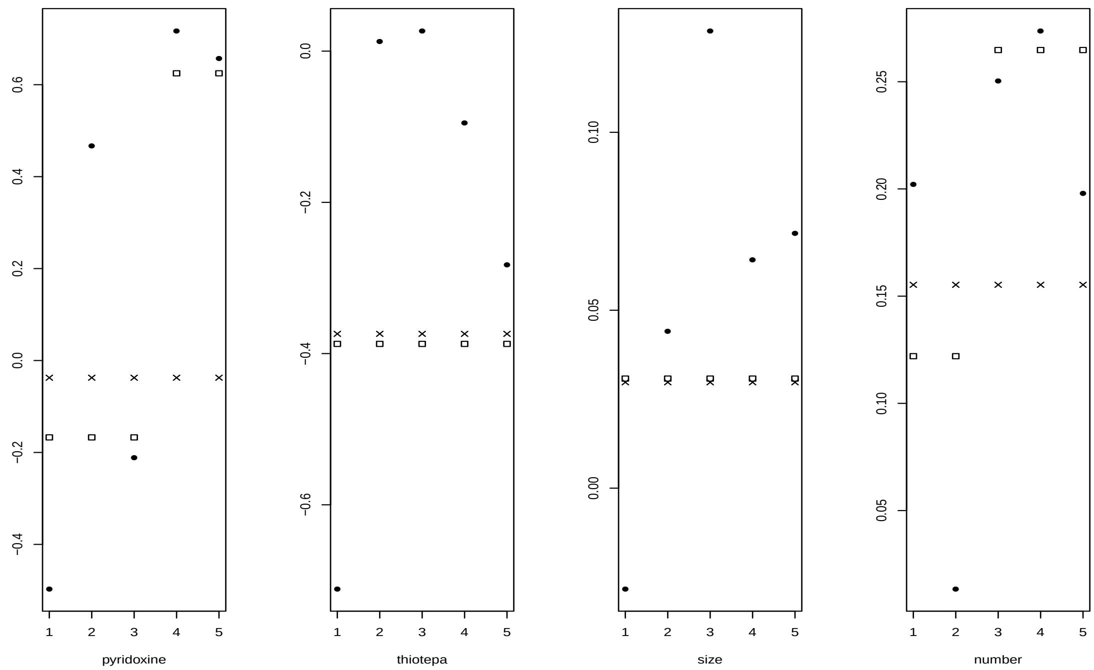

Fig. 1. Estimates for the bladder data in the multiplicative model. The crosses represent the constant estimator, the filled circles the unconstrained estimator and the squares the reweighed lasso estimator.

6. Discussion

In this paper, the Aalen and Cox models were studied to model the effect of covariates on the rate function. However, such models are not essential in our approach. Penalized algorithms could be easily derived for other models such as the accelerated failure time model or the semiparametric transformation model for instance.

Although we have only presented asymptotic theoretical results, the simulation studies show clear evidence that our estimators outperform standard estimators for small sample sizes. Therefore, it would be of great interest to study their finite sample properties. However, such results involve deviation inequalities for non i.i.d. and non martingale empirical processes. To our knowledge, no such results have yet been established in the context of recurrent events.

Another development of the present paper would be to establish results for the estimation of change-point locations and the number of change-points. Such results can be found for the change-point detection in the mean of a gaussian signal in Harchaoui & Lévy-Leduc (2010), for instance.

Appendix: Proofs

Proofs of Lemma 1 to 3 are in Supplementary Material.

A key relation

LEMMA A1. *Under Assumption* 1*, for all* $i = 1, \dots, n$

$$\mathbb{E}\big(dN_i(t) \mid X_i(t), D_i \wedge C_i \geq t, N_i(t) = s-1\big) = Y_i(t)\rho_0(t, s, X_i(t))dt.$$

Decomposition of the least squares criterion in the additive model

The next proposition gives the details of the construction of the partial least squares in the additive model. One has to notice that the processes $Z_n(s)$ introduced below are centered which implies that finding a minimizer of L_n^{PLS} is a natural way of estimating β_0 in model (2).

LEMMA A2. *In the additive event-specific model* (2)*, the partial least squares criterion* (5) *can be rewritten as*

$$L_n^{PLS}(\beta) = \sum_{s=1}^{B} \left\{\beta(s)^\top \mathbf{H}_n(s)\beta(s) - 2\beta(s)^\top \mathbf{H}_n(s)\beta_0(s) - 2Z_n(s)\beta(s)\right\}, \tag{A1}$$

where

$$Z_n(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{B}\int \{X_i(t) - \bar{X}^s(t)\}\mathbf{1}(N_i(t) = s)dM_i^s(t).$$

A technical lemma

LEMMA A3. *Let* $\mathcal{D}[0,\tau]$ *denotes the set of càdlàg functions on* $[0,\tau]$ *and let* $F_n(\cdot)$ *and* $f(T, \delta, X(\cdot), N(\cdot))$ *be two random processes of bounded variation on* $[0,\tau]$*. Suppose that for all* $z \in [0,\tau]$,

$$\mathbb{E}\left[\Big(\int_0^z f(T, \delta, X(t), N(t))dM^s(t)\Big)^2\right] < \infty.$$

We then have the following properties:

(i) If $f(T, \delta, X(\cdot), N(\cdot))$ *is a random variable of bounded variation on* $[0,\tau]$*, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^z f(T_i, \delta_i, X_i(t), N_i(t))dM_i^s(t)$$

converges weakly in $\mathcal{D}[0,\tau]$ *to a centered gaussian process with variance equal to*

$$\mathbb{E}\left[\Big(\int_0^z f(T, \delta, X(t), N(t))dM^s(t)\Big)^2\right].$$

(ii) If $\sup_{t\in[0,\tau]}|F_n(t) - F(t)| = o_{\mathbb{P}}(1)$*, where* $F(\cdot)$ *is a random process on* $[0,\tau]$*, then*

$$\sup_{z\in[0,\tau]}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^z (F_n(t) - F(t))f(T_i, \delta_i, X_i(t), N_i(t))dM_i^s(t)\right\} = o_{\mathbb{P}}(1).$$

Proof of Theorem 1

PROOF OF 1. Let $\Gamma_n^{add}(\beta)$ be the quantity minimized by $\hat{\beta}_{\text{TV}/add}$ and introduce $\Gamma_{add}(\beta) = \sum_{s=1}^{B}\left[\beta(s)^\top \mathbf{H}(s)\beta(s) - 2\boldsymbol{h}(s)\beta(s)\right]$ where

$$\boldsymbol{h}(s) := \int \mathbb{E}\left[\mathbf{1}(N(t) = s)X(t)dN(t)\right] - \int \frac{\mathbb{E}[Y^s(t)X(t)]}{\mathbb{E}[Y^s(t)]}\mathbb{E}[\mathbf{1}(N(t) = s)dN(t)].$$

Using Lemma A1 notice that $\boldsymbol{h}(s) = \beta_0(s)^\top \mathbf{H}(s)$ and consequently, $\text{argmin}_\beta \Gamma_{add} = \beta_0$. Since the criterion to minimize is convex, the convergence in probability of $\hat{\beta}_{\text{TV}/add}$ to β_0 follows from the

pointwise convergence of $\Gamma_n^{add}(\beta)$ towards $\Gamma_{add}(\beta)$. Now write:

$$\begin{aligned}\left|\Gamma_n^{add}(\beta)-\Gamma_{add}(\beta)\right| &\leq \left|L_n^{PLS}(\beta)-\Gamma(\beta)\right| + \frac{\lambda_n}{n}Bp\max_{s,j}|\beta^j(s)-\beta^j(s-1)| \\ &\leq Bp^2\max_{j,k,s}|\beta^j(s)\beta^k(s)(\mathbf{H}_n^{j,k}(s)-\mathbf{H}^{j,k}(s))| + 2Bp\max_{j,s}|\boldsymbol{h}_n^j(s)-\boldsymbol{h}^j(s)||\beta^j(s)| + \frac{\lambda_n}{n}Bp\end{aligned}$$

and the result follows from the law of large number and the fact that $\lambda_n/n \to 0$ as n tends to infinity.

Proof of 2. Define

$$\Lambda_n^{add}(u) = \sum_{s=1}^B u(s)^\top \mathbf{H}_n(s)u(s) - 2\sqrt{n}\sum_{s=1}^B Z_n(s)u(s) + \lambda_n\sum_{j=1}^p\Big(\text{TV}(\beta_0^j+u^j/\sqrt{n}) - \text{TV}(\beta_0^j)\Big)$$

and notice that $\Lambda_n^{add}(u)$ is minimum at $u=\sqrt{n}(\hat\beta_{\text{TV}/add}-\beta_0)$. Write

$$\begin{aligned}\sqrt{n}\sum_{s=1}^B Z_n(s)u(s) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\sum_{s=1}^B\left(X_i(t)-\frac{\mathbb{E}[Y^s(t)X(t)]}{\mathbb{E}[Y^s(t)]}\right)u(s)\mathbf{1}(N_i(t)=s)dM_i^s(t) \\ &- \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\sum_{s=1}^B\left(\bar X^s(t)-\frac{\mathbb{E}[Y^s(t)X(t)]}{\mathbb{E}[Y^s(t)]}\right)u(s)\mathbf{1}(N_i(t)=s)dM_i^s(t).\end{aligned}$$

Let $F_n(t)=\sum_s(\bar X^s(t)-\mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])u(s)$ and $F(t)=0$. F_n has bounded variation and from Lemma A3 (ii), the second term converges to 0 in probability. Now, take $f(T_i,\delta_i,X_i(t),N_i(t))=\sum_s(X_i(t)-\mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])u(s)\mathbf{1}(N_i(t)=s)$ which is also a function of bounded variation. From Lemma A3 (i), the first term converges weakly towards a centered gaussian variable with variance equal to

$$\begin{aligned}&\mathbb{E}\left[\Big(\int_0^\tau\sum_{s=1}^B(X(t)-\mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])u(s)\mathbf{1}(N(t)=s)dM^s(t)\Big)^2\right] \\ &= \sum_{s=1}^B u(s)^\top\mathbb{E}\left[\Big(\int_0^\tau(X(t)-\mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])\mathbf{1}(N(t)=s)dM^s(t)\Big)^{\otimes 2}\right]u(s).\end{aligned}$$

Then, note that $\sum_{s=1}^B u(s)^\top\mathbf{H}_n(s)u(s)$ converges to $\sum_{s=1}^B u(s)^\top\mathbf{H}(s)u(s)$, in probability and $\lambda_n\sum_j\Big(\text{TV}(\beta_0^j+u^j/\sqrt{n})-\text{TV}(\beta_0^j)\Big)/\lambda_0$ converges to

$$\sum_{j=1}^p\sum_{s=2}^B\Big\{|\Delta u^j(s)|\mathbf{1}(\Delta\beta^j(s)=0) + \text{sgn}(\Delta\beta_0^j(s))(\Delta u^j(s))\mathbf{1}(\Delta\beta^j(s)\neq 0)\Big\}.$$

Thus $\Lambda_n^{add}(u)$ converges to $\Lambda_{add}(u)$ in distribution. Since Λ_n^{add} is convex and Λ_{add} has a unique minimum, it follows that $\sqrt{n}(\hat\beta_{\text{TV}/add}-\beta_0)$ converges to $\text{argmin}_u\Lambda_{add}(u)$ in distribution.

Proof of Theorem 2

First define for $l=0,1$ or 2

$$S_n^{(l)}(s,t,\beta) = \frac{1}{n}\sum_{i=1}^n Y_i^s(t)X_i(t)^{\otimes l}\exp(X_i(t)\beta(s)).$$

Following the arguments in example VII.2.7 page 502 of Andersen et al. (1993), it can easily be shown that

$$\sup_{t\in[0,\tau]}|S_n^{(l)}(s,t,\beta_0)-s^{(l)}(s,t,\beta_0)| \xrightarrow[n\to\infty]{\mathbb{P}} 0,\ \forall\, l=0,1,2,$$

using the fact that the covariates process is of bounded variation (in particular, this assumption guarantees that $s^{(l)}(s,t,\beta_0)$ has a countable number of jumps).

PROOF OF 1. Let $\Gamma_n^{mult}(\beta)$ be the quantity minimized by $\hat{\beta}_{\text{TV}/mult}$ and introduce

$$\begin{aligned}\Gamma_{mult}(\beta) &= -\sum_{s=1}^{B}\int \mathbb{E}\left[X(t)\beta(s)Y^s(t)dN(t)\right] + \sum_{s=1}^{B}\int \log(s^{(0)}(s,t,\beta))\mathbb{E}\left[Y^s(t)dN(t)\right] \\ &= -\sum_{s=1}^{B}\int \alpha_0(t,s)\left(\beta(s)^\top s^{(1)}(s,t,\beta_0) - \log(s^{(0)}(s,t,\beta))s^{(0)}(s,t,\beta_0)\right)dt,\end{aligned}$$

where the last equality follows from Lemma A1. From similar arguments as in proof 1. of Theorem 1 and the uniform convergence with respect to t of $S_n^{(0)}(s,t,\beta_0)$ towards $s^{(0)}(s,t,\beta_0)$, we get the pointwise convergence in probability of $\Gamma_n^{mult}(\beta)$ to $\Gamma_{mult}(\beta)$. Then, the consistency of $\hat{\beta}_{\text{TV}/mult}$ follows from the convexity of $\Gamma_n^{mult}(\beta)$ and the fact that $\operatorname{argmin}_\beta \Gamma_{mult}(\beta) = \beta_0$.

PROOF OF 2. Consider the convex function

$$\Lambda_n^{mult}(u) = n\Gamma_n(\beta_0 + u/\sqrt{n}) - n\Gamma_n(\beta_0) + \lambda_n \sum_{j=1}^{p}\left(\text{TV}(\beta_0^j + u^j/\sqrt{n}) - \text{TV}(\beta_0^j)\right)$$

which is minimum at $u = \sqrt{n}(\hat{\beta}_{\text{TV}/mult} - \beta_0)$. Then from a Taylor expansion, one gets

$$\begin{aligned}\Lambda_n^{mult}(u) = &-\frac{\sqrt{n}}{n}\sum_{s=1}^{B}\sum_{i=1}^{n}\int \left(X_i(t) - \mathbf{E}_n(s,t,\beta_0)\right)Y_i^s(t)dN_i(t)u(s) \\ &+ \frac{1}{2n}\sum_{s=1}^{B}u(s)^\top \sum_{i=1}^{n}\int \mathbf{V}_n(s,t,\beta_0)Y_i^s(t)dN_i(t)u(s) + \lambda_n\sum_{j=1}^{p}\left(\text{TV}(\beta_0^j + u^j/\sqrt{n}) - \text{TV}(\beta_0^j)\right) + o_{\mathbb{P}}(1),\end{aligned}$$

where

$$\mathbf{E}_n(s,t,\beta) = \frac{S_n^{(1)}(s,t,\beta)}{S_n^{(0)}(s,t,\beta)}, \quad \mathbf{V}_n(s,t,\beta) = \frac{S_n^{(2)}(s,t,\beta)}{S_n^{(0)}(s,t,\beta)} - \mathbf{E}_n(s,t,\beta)^{\otimes 2}.$$

The uniform convergence with respect to t of $S_n^{(0)}(s,t,\beta)$ and $S_n^{(2)}(s,t,\beta)$ towards $s^{(0)}(s,t,\beta_0)$ and $s^{(2)}(s,t,\beta_0)$ respectively and the law of large number give the convergence in probability of the term

$$\frac{1}{2n}\sum_{s=1}^{B}u(s)^\top\sum_{i=1}^{n}\int \mathbf{V}_n(s,t,\beta_0)Y_i^s(t)dN_i(t)u(s)$$

towards

$$\frac{1}{2}\sum_{s=1}^{B}u(s)^\top \int \mathbf{v}(s,t,\beta_0)\mathbb{E}[Y^s(t)dN(t)]u(s).$$

Notice that

$$\sum_{i=1}^{n}\left(X_i(t) - \mathbf{E}_n(s,t,\beta_0)\right)Y_i^s(t)\alpha_0(t,s)\exp(X(t)\beta_0(t))dt = 0$$

in order to rewrite the first term of $\Lambda_n^{mult}(u)$ as

$$-\frac{\sqrt{n}}{n}\sum_{s=1}^{B}\sum_{i=1}^{n}\int \left(X_i(t) - \mathbf{E}_n(s,t,\beta_0)\right)u(s)Y_i^s(t)dM_i^s(t).$$

From Lemma 3, the same kind of arguments as in the proof of Theorem 1 can be applied to conclude the proof.

References

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory (Proc. Sixth Internat. Conf., Wisła, 1978)*, vol. 2 of *Lecture Notes in Statist.* New York: Springer, pp. 1–25.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. New York: Springer-Verlag.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–1120.

Bouaziz, O., Comte, F. & Guilloux, A. (2013). Nonparametric estimation of the intensity function of a recurrent event process. *Statistica Sinica* **23**, 635–665.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage 1 bladder tumors: comparison of placebo, pyridoxine, and topical thiotepa. *Bladder Tumors and Others Topics in Urological Oncology* , (Edited by M. Pavone–Macaluso, P. H. Smith, and F. Edsmyn), 363–370.

Cai, J. & Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Anal.* **10**, 121–138.

Candès, E. J., Wakin, M. B. & Boyd, S. P. (2008). Enhancing sparsity by reweighted l_1 minimization. *J. Fourier Anal. Appl.* **14**, 877–905.

Dauxois, J.-Y. & Sencey, S. (2009). Non-parametric tests for recurrent events under competing risks. *Scand. J. Stat.* **36**, 649–670.

Gaiffas, S. & Guilloux, A. (2012). High-dimensional additive hazards models and the lasso. *Electronic Journal of Statistics* **6**, 522–546.

Goossens, M. E., Buntinx, F., Joniau, S., Ackaert, K., Ameye, F., Billiet, I., Braeckman, J., Breugelmans, A., Darras, J., Dilen, K., Goeman, L., Kellen, E., Tombal, B., Van Bruwaene, S., Van Cleyenbreuge, B., Van der Aa, F., Vekemans, K., Van Poppel, H. & Zeegers, M. (2012). Designing the selenium and bladder cancer trial (seleblat), a phase lll randomized chemoprevention study with selenium on recurrence of bladder cancer in belgium. *BMC Urology* .

Harchaoui, Z. & Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105**, 1480–1493.

Huang, Y. & Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Anal.* **9**, 293–303.

Huffer, F. W. & McKeague, I. (1991). Weighted least squares estimation for aalen's additive risk model. *Journal of the American Statistical Association* **86**, 114–129.

Kelly, P. & Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Stat. Med.* **19**.

Lawless, J. F. & Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–168.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62**, 711–730.

Lin, D. Y., Wei, L. J. & Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.

Lin, D. Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

Martinussen, T. & Scheike, T. H. (2009a). The additive hazards model with high-dimensional regressors. *Lifetime Data Anal.* **15**, 330–342.

Martinussen, T. & Scheike, T. H. (2009b). Covariate selection for the semiparametric additive risk model. *Scand. J. Stat.* **36**, 602–619.

McKeague, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, vol. 80 of *Contemp. Math.* Providence, RI: Amer. Math. Soc., pp. 139–152.

Prentice, R. L., Williams, B. J. & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379.

Scheike, T. H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. *Lifetime Data Anal.* **8**, 247–262.

Sun, L., Park, D.-H. & Sun, J. (2006). The additive hazards model for recurrent gap times. *Statist. Sinica* **16**, 919–932.

Tanaka, T., Miyazawa, K., Tsukamoto, T., Kuno, T. & Suzuki, K. (2011). Pathobiology and chemoprevention of bladder cancer. *Journal of Oncology* .

WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065–1073.

ZENG, D. & CAI, J. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97**, 699–712.

SUPPLEMENTARY MATERIAL

Supplementary material includes a description of the algorithms, extended simulation study and additional analysis on the bladder tumour data of Byar (1980). It also contains proofs of Proposition 2 and Lemma 3.